June 13, 2024

CUSTOMER ANALYSIS REPORT

Presented To

Long

Presented By

Dat Duong

duongleminhdat20040512@gmail.com

437 255 2999

Table of Contents

Overview

What is Customer360?

Customer360 is an integrated approach to analyzing and utilizing customer data from multiple dimensions to generate a comprehensive perspective of customer interactions and behaviors. This technique aggregates data into four categories: Demographics Data, Transaction Data, Behavioral Data, Interaction Data.

- **Demographics Data:** includes attributes like age, gender, income, and other characteristics that help define the customer profile.
- **Transaction Data:** cover details about customer purchases, payment methods, order history, and usage patterns.
- **Behavioral Data:** reflects customer preferences, needs, desires, and opinions, offering insights into customer motivation and tendencies.
- **Interaction Data:** encompasses all touchpoints between the customer and the business, such as email, social media interactions, call center notes, and web click streams.

Figure 1. Customer360

Why Customer360 is Important?

When businesses understand their customers, they can deliver superior customer experiences while securing their own competitive advantage. Here are some key reasons why Customer360 is important:

- **Exceptional Customer Experience:** by understanding customer preferences and behaviors, businesses can tailor their interactions to meet individual needs, enhancing satisfaction and loyalty. Personalized experiences lead to increased customer retention and repeat visits.
- **Reduced Customer Frustration:** by accurately managing inventory and aligning marketing campaigns with product availability, businesses can avoid customer frustration. Preventing situations where customers are directed to unavailable products ensures a smoother shopping experience.

- **Improved Data Quality for Analytics:** high quality, integrated customer data enhances the accuracy of analytics and machine learning models. Precise data inputs lead to better decision-making and more effective business strategies.

RFM Model

RFM is an effective customer segmentation technique where it will be very helpful for marketers, to make strategic choices in the business. It engages marketers to rapidly distinguish and segment customers into similar clusters and target them with separated and personalized promoting methodologies. This in turn makes strides in customer engagement and retention.

- **Recency (R):** assesses how recently a customer has transacted with your brand. Customers who have made a purchase recently are more likely to do so again in the near future.
- **Frequency (F):** measures how often a customer engages with your brand. Those who interact frequently are more likely to continue doing so.
- **Monetary (M):** represents the amount of money a customer has spent on your products or services. High spenders are likely to remain big spenders.

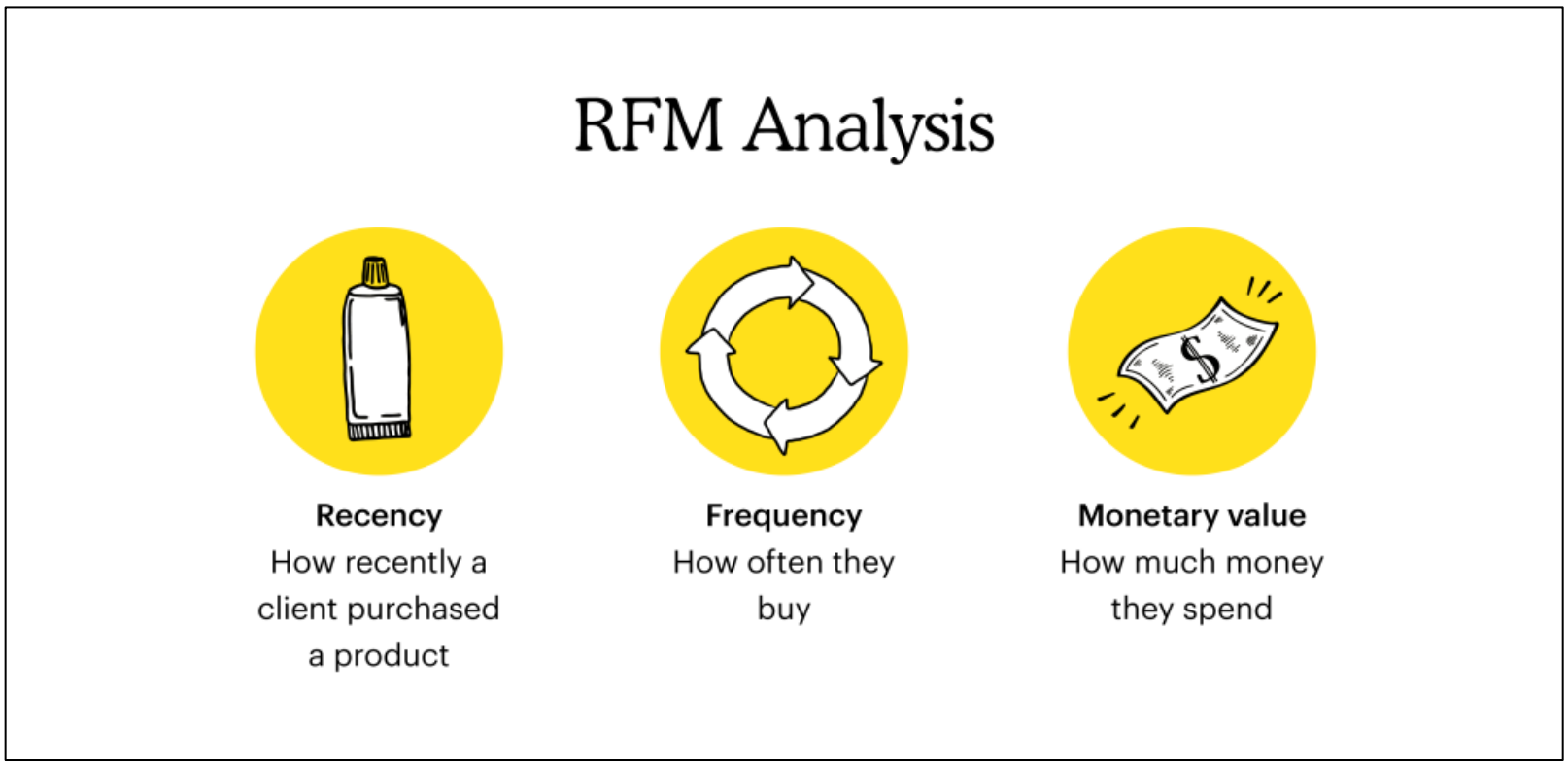


Figure 2. RFM Model

IQR Method

The interquartile range (IQR) is a measure of statistical dispersion, which is the spread of the middle 50% of a data set. It is calculated by subtracting the first quartile (Q1) from the third quartile (Q3).

- First Quartile (Q1): median of the lower half of the data
- Second Quartile (Q2): median of the dataset
- Third Quartile (Q3): median of the upper half of the data

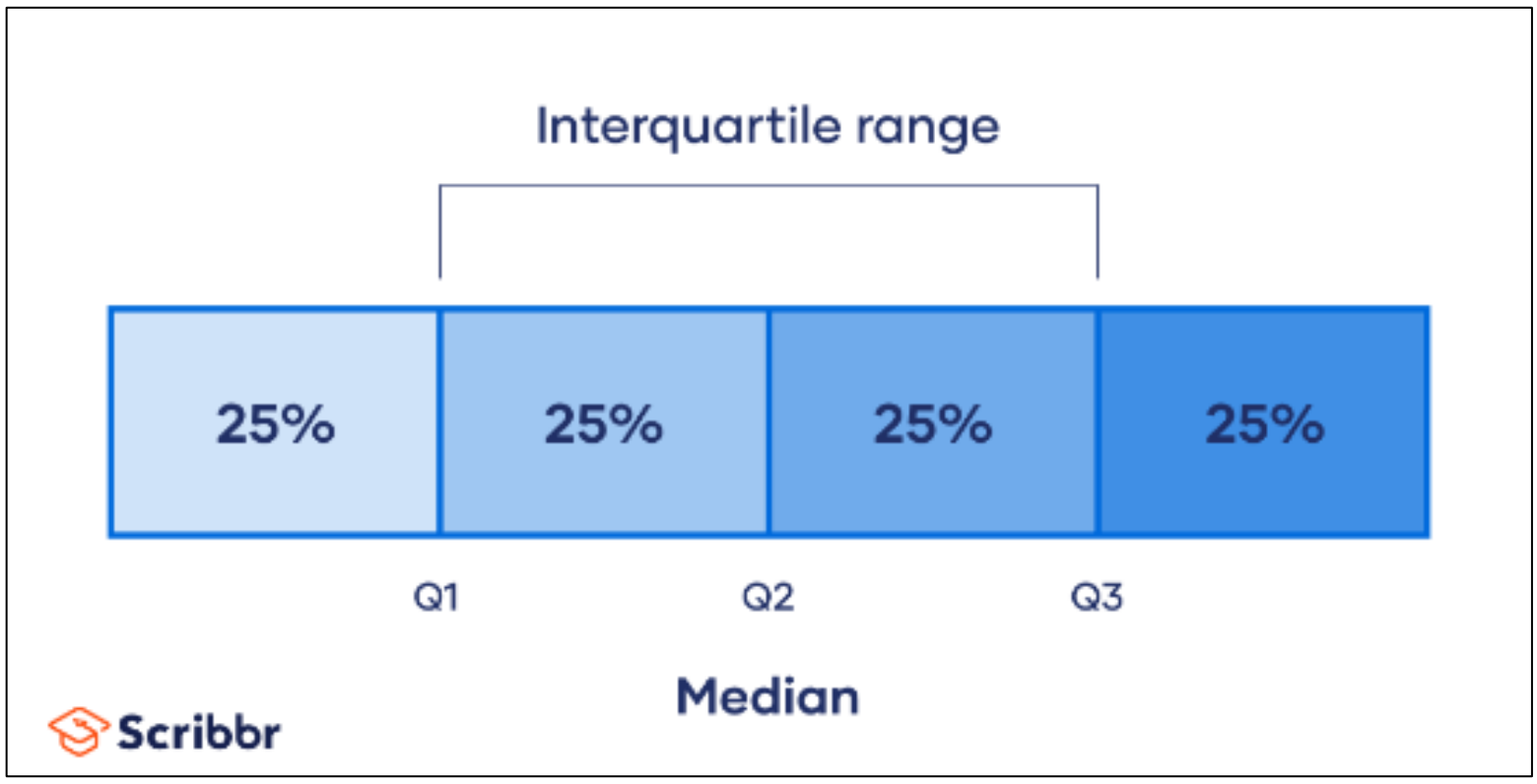


Figure 3. Interquartile Range

BCG Matrix

Also known as Boston Consulting Group Matrix, is a strategic business tool used to evaluate a company's product portfolio and inform decision-making regarding investment, development, and discontinuation of products.

- **Stars:** High growth, high market share products that require significant investment but have the potential for high returns.
- **Cash Cows:** Low growth, high market share products that generate steady cash flow with minimum investment.
- **Question Marks:** High growth, low market share products that require substantial investment to increase market share and determine their future potential.
- **Dogs:** Low growth, low market share products that may drain resources and are often candidates for divestiture.

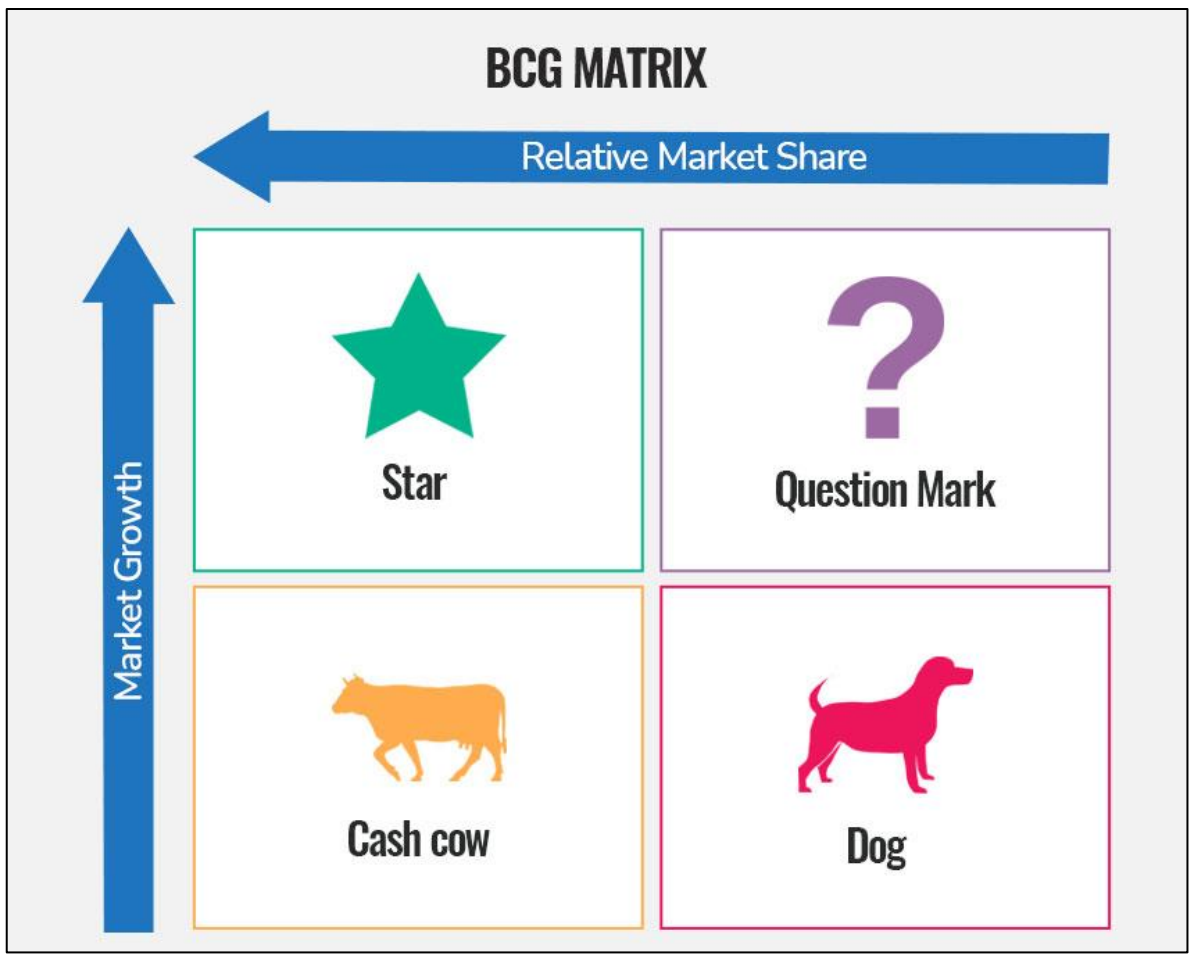


Figure 4. BCG Matrix

EDA

Data

- Customer_Registered Table:

Column name	Data type
ID	VARCHAR
Contract	VARCHAR
LocationID	VARCHAR
BranchCode	VARCHAR
Status	VARCHAR
Created_date	DATETIME
Stopdate	DATETIME

	ID	Contract	LocationID	BranchCode	Status	created_date	stopdate
1	1	SGDN00215	8	1	0	2011-11-25	2012-01-05
2	2	SGDN00214	8	1	0	2012-06-14	<null>
3	3	SGD374348	8	1	0	2012-11-01	<null>
4	4	SGD022064	8	1	2	2011-06-22	2013-05-29
5	5	SGD041015	8	5	2	2011-12-17	2014-11-11
6	6	SGDN00211	<null>	<null>	2	2015-06-09	2015-09-09
7	7	SGD374348	8	1	3	2012-11-26	2012-12-13
8	8	SGDN00013	0	0	1	2017-12-11	<null>
9	9	BEAAA1809	75	0	1	2022-04-04	<null>
10	10	BNAAA4298	241	2	1	2022-04-19	<null>
11	11	SGAAH8487	8	3	1	2022-04-25	<null>
12	12	SGDN00017	0	0	2	2017-12-04	2018-04-27
13	13	SGDN00017	0	0	2	2017-12-04	2020-02-12
14	14	SGB000137	0	0	2	2017-12-05	2017-12-05
15	15	SGDN00017	0	0	2	2017-12-05	2020-06-26
16	16	SGDN00017	0	0	2	2017-12-05	2017-12-06
17	17	SGH068462	0	0	2	2017-12-06	2017-12-06
18	18	SGH068460	0	0	2	2017-12-06	2017-12-06
19	19	SGDN00017	0	0	2	2017-12-06	2018-01-10
20	20	SGH068462	0	0	2	2017-12-06	2017-12-07

Figure 5. Sample Customer_Registered Data

- Customer_Transaction Table:

Column name	Data type
Transaction_ID	VARCHAR
CustomerID	VARCHAR
Purchase_Date	DATETIME
GMV	BIGINT

	Transaction_ID	CustomerID	Purchase_Date	GMV
1	0	1327813	2022/06/01	95000
2	1	1157830	2022/06/01	75000
3	2	873915	2022/07/01	95000
4	3	3505071	2022/07/01	90000
5	4	2930918	2022/07/01	109091
6	5	899882	2022/06/01	105000
7	6	2248818	2022/06/01	75000
8	7	3331485	2022/06/01	90000
9	8	3497579	2022/06/01	60000
10	9	2636994	2022/07/01	75000

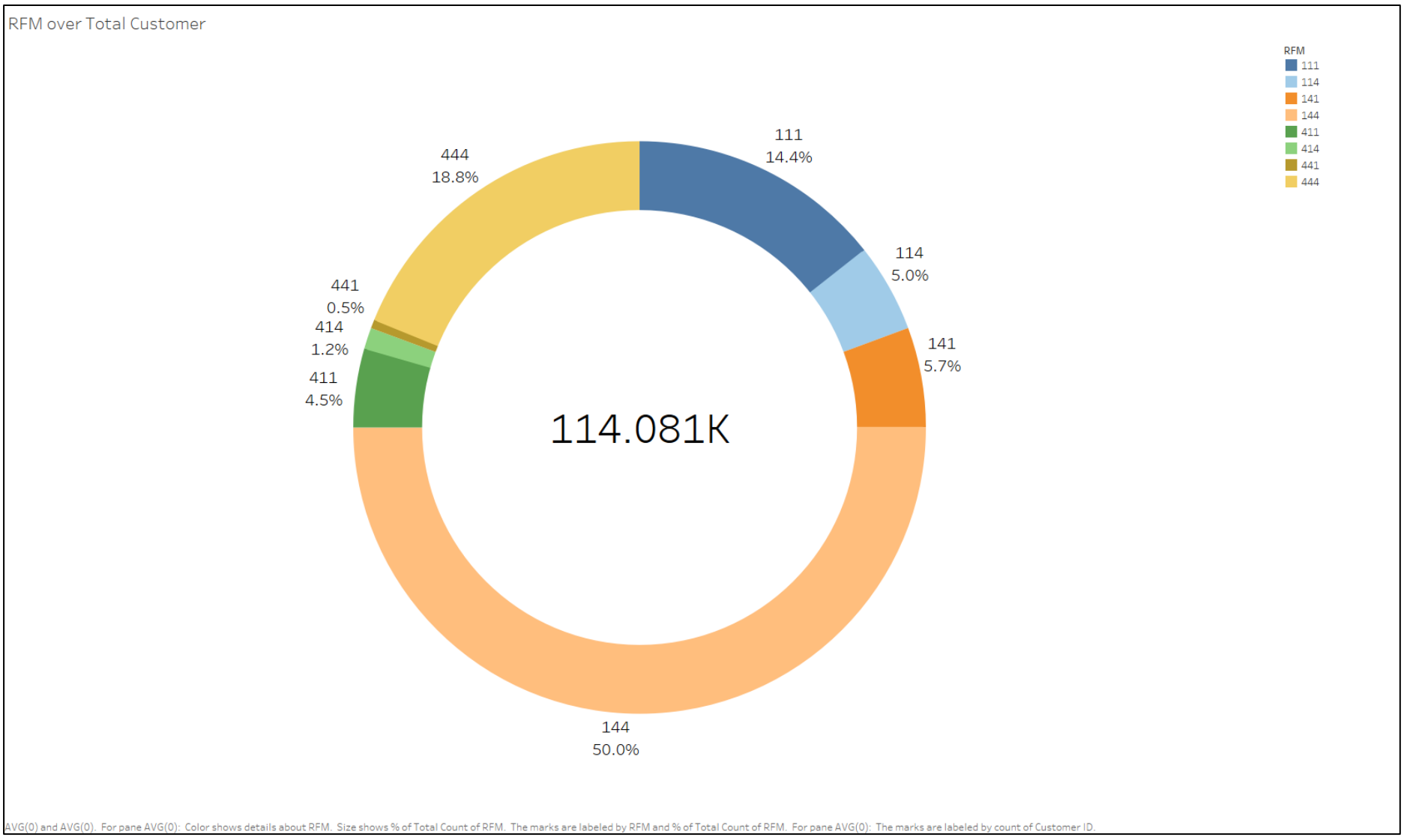
Figure 6. Sample Customer_Transaction Data

Explore & Analyze

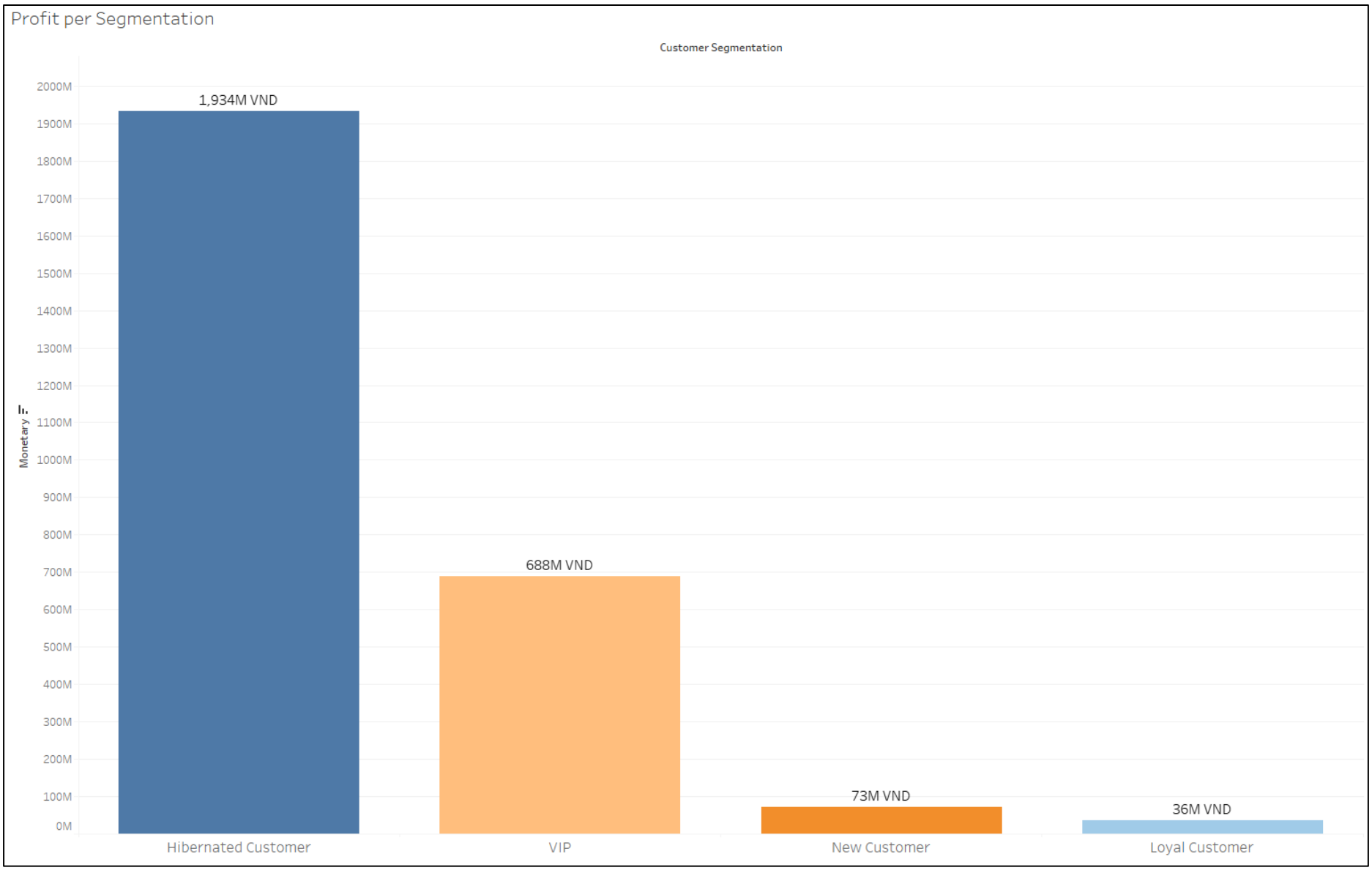
Calculate RFM indices:

- Take '2022/09/01' as the beginning day.
- **Recency**: the difference between '2022-09-01' and the nearest clients logged in.
- **Frequency**: $\frac{2022/09/01 \ - \textit{the most recent purchase customer made}}{2022/09/01 \ - \ \textit{contract's creation date}}$
- **Monetary Value**: $\frac{\textit{total money spent}}{\#\ of\ purchases}$
- By using the IQR method to determine the RFM indices, we're able to categorize customers into 4 main segmentations (VIP, Loyal Customer, New Customer, Hibernated Customer).

Segment	**Characteristics**	**RFM**
VIP	Customers who frequently use the service and have made significant expenditures recently.	444, 443, 442, 434, 433, 432, 344, 343, 333, 334
Loyal Customer	Customers who use the service infrequently but have made significant expenditures recently.	441, 431, 424, 423, 414, 413, 342, 332, 323, 324, 331, 243, 244, 234
New Customer	Customers who frequently use the service but do not spend much.	422, 421, 412, 411, 311, 312, 313, 314, 321, 322, 341
Hibernated Customer	Customers who spend little and do not visit very often.	111, 121, 122, 123, 124, 131, 132, 133, 134, 141, 142, 143, 144

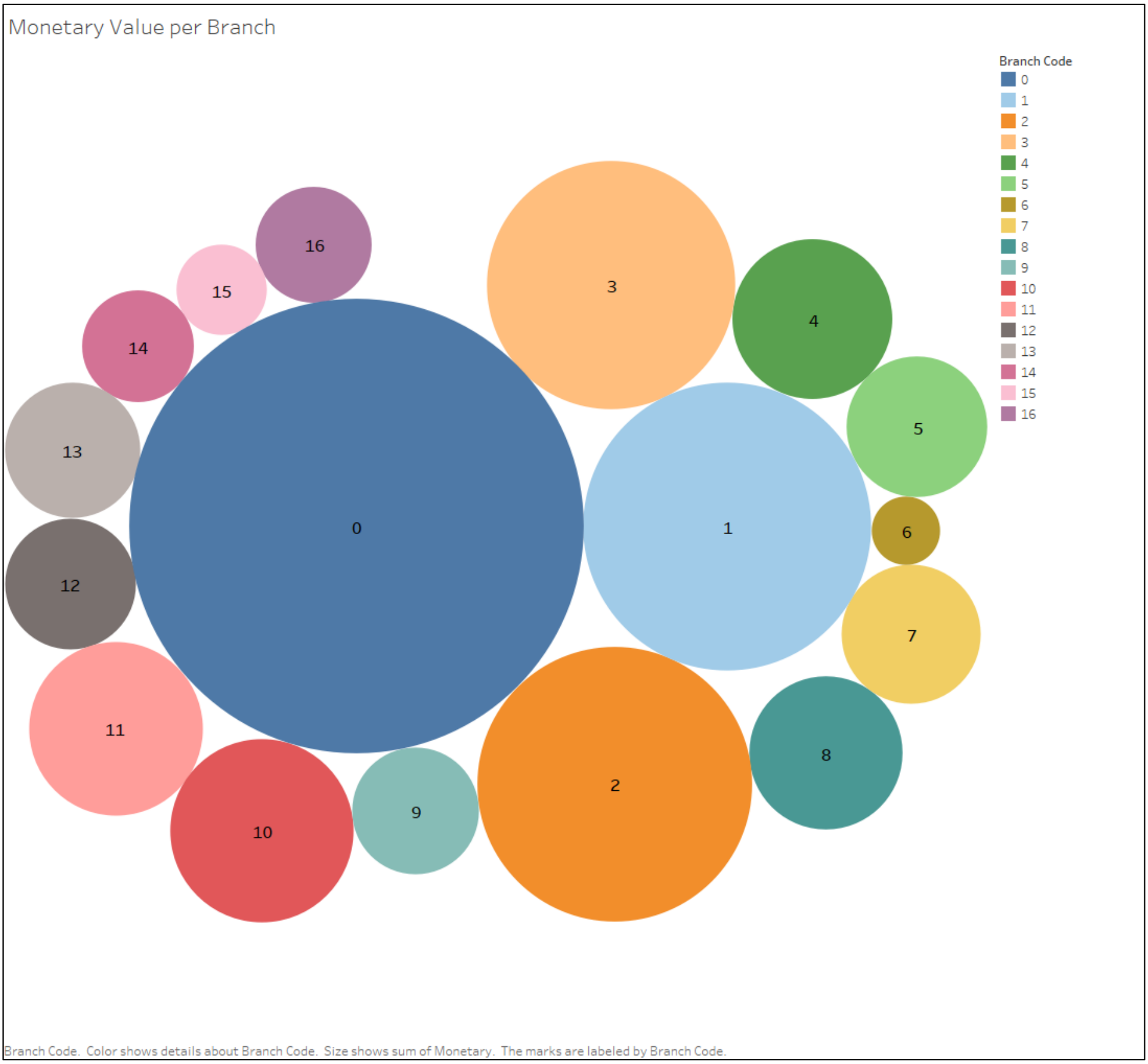


The donut chart reveals the distribution of customer loyalty based on RFM indices among a total of 114.081K customers. The VIP Customers making up 18.8% with an RFM score of 444 are the most loyal and valuable, frequently using the service and making significant expenditures recently. New Customers constituting 5% with an RFM score of 114, frequently use the service but do not spend much yet, showing potential for increased loyalty. Hibernated Customers including 4.5% with an RFM score of 411 and the dominant of 144, showing low engagement and loyalty. To maximize customer loyalty and business growth, personalized marketing should be used to focus on retention and reward programs for VIP and Loyal Customers, engagement tactics for New Customers, and reactivation efforts for Hibernated Customers.



The bar graph illustrates the profitability of different customer segments. Notably, Hibernated Customers generate the highest profit, contributing significantly to overall revenue (1.934 million VND). Following are VIP Customers, who contribute 688 million VND. In contrast, New Customers and Loyal Customers yield lower profits (73 million VND and 36 million VND respectively). The reason for this huge difference is due to the majority of customers are hibernated. To enhance profitability:

- Re-engage Hibernated Customers: prioritize efforts to reactivate this segment, as they represent the largest revenue source.
- Convert New and Loyal Customers: aim to elevate them to VIP status and prevent them to become hibernated.



The chart represents the monetary value per branch, with each bubble's size corresponding to the respective branch's monetary value. Branch Code 0 stands out with the largest bubble, indicating the highest monetary value. Following are branch 1, 2, and 3 also have significant monetary values. In contrast, other branches are relatively smaller and need to be paid more attention.

SQL Code

Calculate RFM indices and store results in rfm table:

```sql
# Calculate recency, frequency, monetary:
CREATE TABLE IF NOT EXISTS cus_rfm AS (SELECT CustomerID,
                                              Contract,
                                              LocationID,
                                              BranchCode,
                                              Status,
                                              DATEDIFF('2022-09-01', MAX(Purchase_Date)) AS recency,
                                              ROUND(1.00 * COUNT(DISTINCT Purchase_Date) /
                                                    TIMESTAMPDIFF(YEAR, created_date, '2022/09/01'),
                                                    2)                                    AS frequency,
                                              ROUND(1.00 * SUM(GMV) /
                                                    TIMESTAMPDIFF(YEAR, created_date, '2022/09/01'),
                                                    2)                                    AS monetary
                                       FROM customer_transaction CT
                                                JOIN customer_registered CR
                                                     ON CT.CustomerID = CR.ID
                                       WHERE CustomerID != 0
                                       GROUP BY CustomerID, created_date);

CREATE TABLE IF NOT EXISTS rfm_calculation AS (SELECT *,
                                                      ROW_NUMBER() over (ORDER BY recency)   AS rn_recency,
                                                      ROW_NUMBER() over (ORDER BY frequency) AS rn_frequency,
                                                      ROW_NUMBER() over (ORDER BY monetary)  AS rn_monetary
                                               FROM cus_rfm);
```

```sql
CREATE TABLE IF NOT EXISTS rfm AS (SELECT *,
      CASE
          WHEN rn_recency >= (SELECT MIN(rn_recency) FROM rfm_calculation) AND
               rn_recency < (SELECT COUNT(rn_recency) * 0.25 FROM rfm_calculation)
              THEN '4'
          WHEN rn_recency >=
               (SELECT COUNT(rn_recency * 0.25) FROM rfm_calculation) AND
               rn_recency < (SELECT COUNT(rn_recency) * 0.5 FROM rfm_calculation)
              THEN '3'
          WHEN rn_recency >=
               (SELECT COUNT(rn_recency * 0.5) FROM rfm_calculation) AND
               rn_recency < (SELECT COUNT(rn_recency) * 0.75 FROM rfm_calculation)
              THEN '2'
          ELSE '1'
          END AS R,
      CASE
          WHEN rn_frequency >= (SELECT MIN(rn_frequency) FROM rfm_calculation) AND
               rn_frequency <
               (SELECT COUNT(rn_frequency) * 0.25 FROM rfm_calculation) THEN '1'
          WHEN rn_frequency >=
               (SELECT COUNT(rn_frequency * 0.25) FROM rfm_calculation) AND
               rn_frequency <
               (SELECT COUNT(rn_frequency) * 0.5 FROM rfm_calculation) THEN '2'
          WHEN rn_frequency >=
               (SELECT COUNT(rn_frequency * 0.5) FROM rfm_calculation) AND
               rn_frequency <
               (SELECT COUNT(rn_frequency) * 0.75 FROM rfm_calculation) THEN '3'
          ELSE '4'
          END AS F,
      CASE
          WHEN rn_monetary >= (SELECT MIN(rn_monetary) FROM rfm_calculation) AND
               rn_monetary < (SELECT COUNT(rn_monetary) * 0.25 FROM rfm_calculation)
              THEN '1'
          WHEN rn_monetary >=
               (SELECT COUNT(rn_monetary * 0.25) FROM rfm_calculation) AND
               rn_monetary < (SELECT COUNT(rn_monetary) * 0.5 FROM rfm_calculation)
              THEN '2'
          WHEN rn_monetary >=
               (SELECT COUNT(rn_monetary * 0.5) FROM rfm_calculation) AND
               rn_monetary < (SELECT COUNT(rn_monetary) * 0.75 FROM rfm_calculation)
              THEN '3'
          ELSE '4'
          END AS M
FROM rfm_calculation
GROUP BY CustomerID, rn_recency, rn_frequency, rn_monetary, recency, monetary,
         frequency);
```

Categorize customers into 4 segments:

```sql
SELECT *,
       CONCAT(R, F, M) AS RFM,
       CASE
           WHEN CONCAT(R, F, M) IN ('444', '443', '442', '434', '433', '432', '344', '343', '333', '334') THEN 'VIP'
           WHEN CONCAT(R, F, M) IN
                ('441', '431', '424', '423', '414', '413', '342', '332', '323', '324', '331', '243', '244', '234')
               THEN 'Loyal Customer'
           WHEN CONCAT(R, F, M) IN ('422', '421', '412', '411', '311', '312', '313', '314', '321', '322', '341')
               THEN 'New Customer'
           ELSE 'Hibernated Customer'
           END         AS customer_segmentation
FROM rfm;
```